AWARD NUMBER: W81XWH-13-1-0214

TITLE: Blood-based biomarkers of early-onset breast cancer

PRINCIPAL INVESTIGATOR: Nasim Ahmadiyeh

CONTRACTING ORGANIZATION: University of California, San Francisco
San Francisco, CA 94143

REPORT DATE: October 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2015	Annual	30 Sep 2014 - 29 Sep 2015

4. TITLE AND SUBTITLE
Blood-based biomarkers of early-onset breast cancer

5a. CONTRACT NUMBER
W81XWH-13-1-0214

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Nasim Ahmadiyeh, Laura Esserman, Allan Balmain, David Quigley

E-Mail:nasim.ahmadiyeh@ucsfmedctr.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Regents of the University
of California
3333 California Street, Suite
315
San Francisco, CA, 94143

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Women with early-onset breast cancer are thought to have a higher contribution of inherited risk than those forming sporadic cancers at later ages. This inherited susceptibility to breast cancer might manifest as differences in gene expression patterns within key oncogenic pathways. While the normal breast is the ideal tissue in which to study this phenomenon, gene expression profiling of blood lymphocytes has been successfully used as a proxy in a variety of diseases including breast cancer. We investigated the gene expression profile of untransformed blood lymphocytes in order to discover gene expression (mRNA and miRNA) signatures which can differentiate BRCA 1/2 negative women with a personal history of early-onset breast cancer and family history of breast cancer (n=51) from asymptomatic aged-matched women without a personal history of cancer or family history of breast cancer (n=50). Using adaboost computer learning algorithm which discretizes the data, and also using logistic elastic net – a form of linear regression - we were unable to build a classifier that could accurately differentiate cases from controls, at any level higher than that already available by history-based risk assessment algorithms.

15. SUBJECT TERMS
biomarkers, early-onset breast cancer, expression profiling, risk-assessment, breast cancer, genomics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	10	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. INTRODUCTION:

The aim of this project, "Blood-based biomarkers of early-onset breast cancer" was to develop a gene-expression signature from peripheral blood, which can accurately predict an individual's risk of developing early-onset breast cancer. Women who are diagnosed with breast cancer before age 40 are more likely to die from their disease than postmenopausal women diagnosed with the same stage breast cancer. This has led many to believe that there is a strong biological/inherited basis to the breast cancer that manifests in younger women. We sought to capture this genetic variation at the level of gene expression differences in peripheral lymphocytes. We compared both mRNA and miRNA profiling of total RNA extracted from peripheral lymphocytes of a cohort of women (n=50) who developed breast cancer by age 45, with a strong family history of breast cancer, but who were BRCA1/2 negative to those of asymptomatic women presenting for screening mammogram with no family history of breast cancer (n=51). The women with early-onset breast cancer were disease and treatment free for at least 6 months at time of blood donation. Cases and controls were age matched to age at blood donation.

2. KEYWORDS: biomarkers, early-onset breast cancer, expression profiling, risk-assessment, breast cancer, genomics

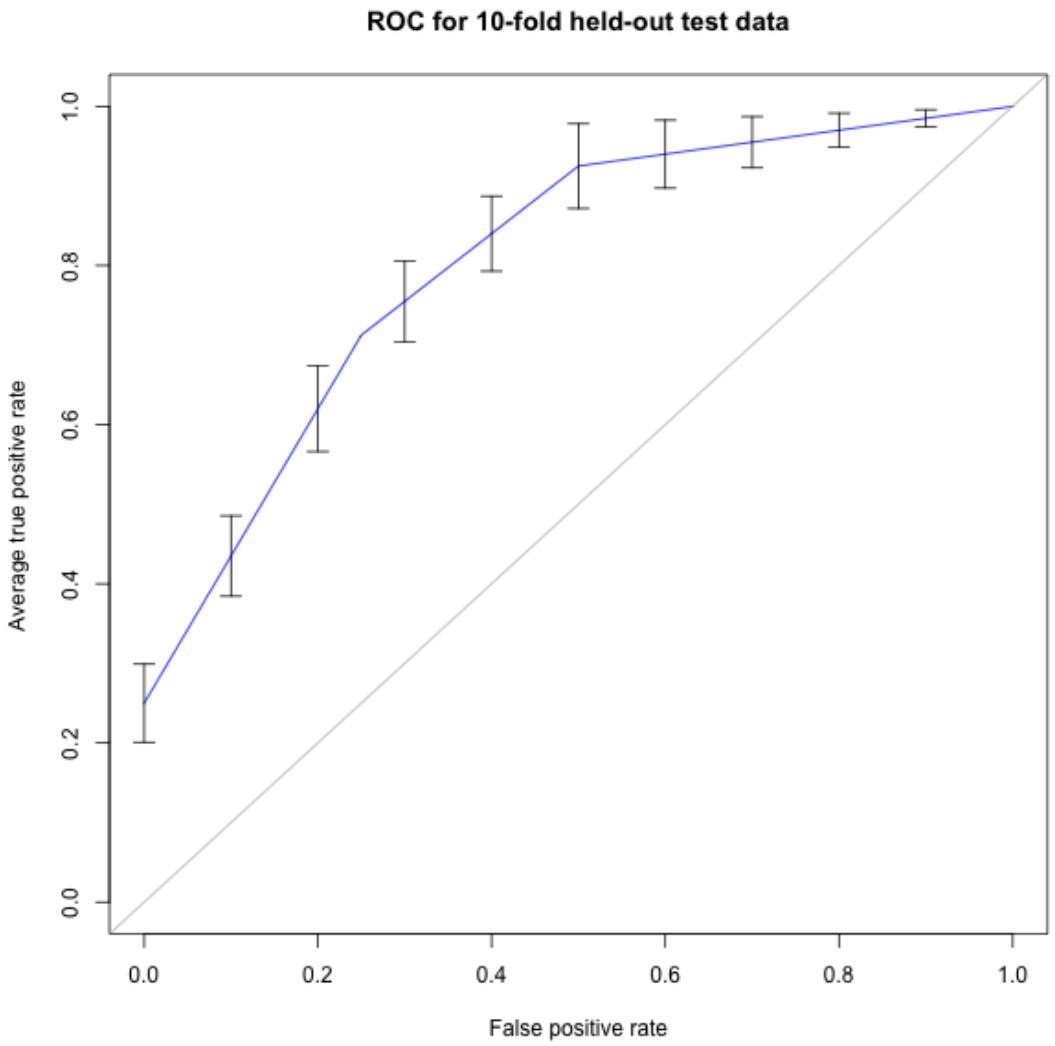
3. ACCOMPLISHMENTS:

Major goals of the project and its accomplishments:

Specific Aim 1: To identify gene expression signatures in blood, which can differentiate known BRCA1/2 negative women with early-onset breast cancer from age-matched asymptomatic women with no history of breast cancer.

Total RNA was extracted from buffy coat using Trizol extraction (Life Technologies), linear acrylamide aided precipitation (ARESCO Inc), and clean-up using a modification to the Qiagen RNEasy Min-Elute cleanup kit in order to preserve the miRNA fraction. These were quantified on nanodrop and run on the bioanalyzer to ensure integrity of RNA. In all, 41 out of 50 cases and 44 out of 51 controls had RNA quality meeting criteria for processing by Affymetrix whole transcript array.

We then ran Affymetrix Whole Transcript Human Arrays and Taqman OpenArray Human miRNA in core facilities. We then analyzed the Affymetrix data and the miRNA data, separately then together. All analysis was primarily performed by David Quigley. He utilized the adaboost machine learning algorithm to build a classifier for differentiating cases from controls off discretized data. The first pass analysis demonstrated a 35 gene signature that differentiated cases from controls at an accuracy of 73%, sensitivity of 85% and specificity of 63%. See ROC curve below.



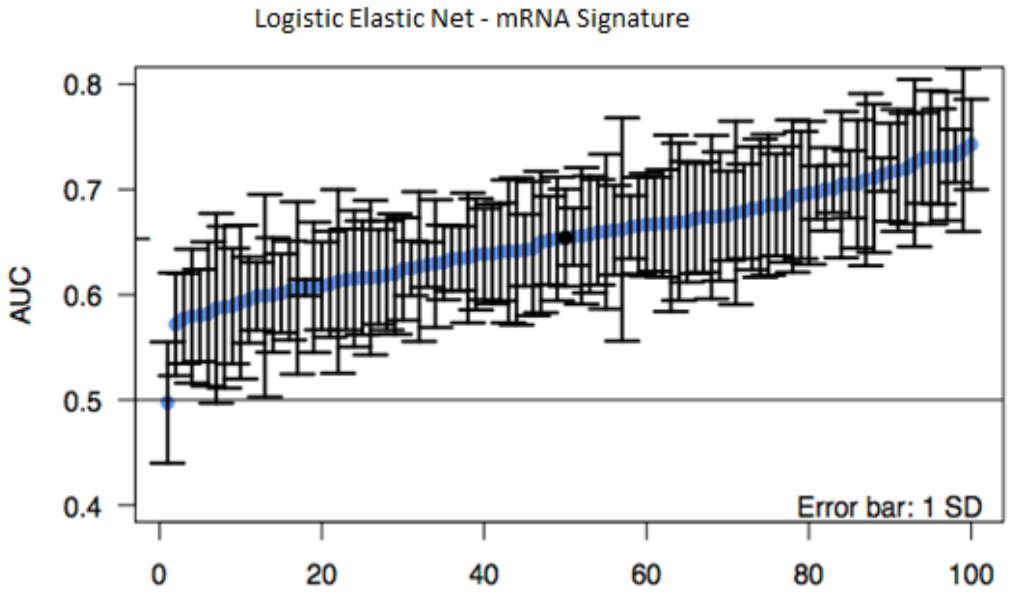
The same approach was used to try to identify a miRNA signature which could reliably differentiate early onset breast cancer cases from controls. No statistically significant signal distinguishing cases from controls was found after performing a cross-validated test using the same adaboost machine learning algorithm described above. Next, a combined mRNA and miRNA signature was attempted, essentially by performing join analysis of the miRNA and mRNA data, again using the adaboost machine learning algorithm. The addition of the miRNA data did not increase the discriminatory power of the classifier produced from mRNA data alone.

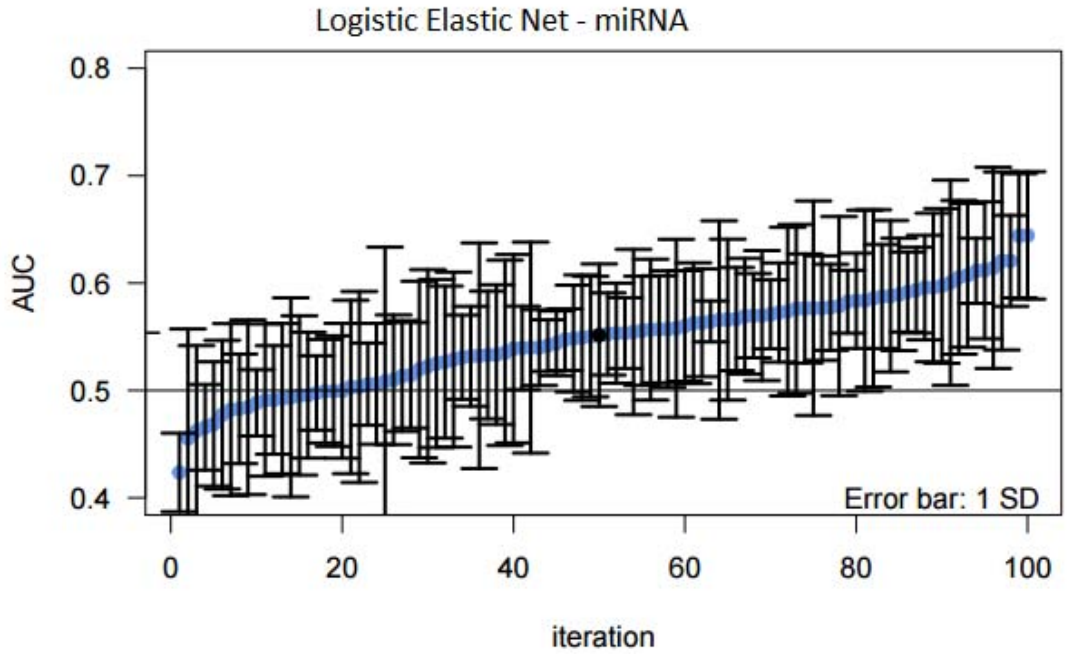
Unfortunately, about a year and a half into the project, we discovered an error in the initial analysis. When applying the adaboost algorithm to the discretized data for mRNA gene signature, the samples used to "train" were also included in the final samples used to "test" the algorithm. This introduced bias by allowing some of the knowledge of the full dataset to leak into the selection features for the individual cross-validation slices. The more correct approach calculates the features to use for each fold individually, without any knowledge of test data.

Once this was discovered, we have spent the remainder of this year trying to re-analyze the data in such a way as to try to derive meaningful result from the cohort. David Quigley turned to

elastic net. Elastic net is a regularized form of linear regression, designed to estimate the parameters that fit a linear function where there may be a large number of independent variables while penalizing overly-complex models. When applying logistic elastic net to our Affymetrix mRNA dataset, there was only 65% accuracy in correctly classifying cases from controls. This is about the same degree of accuracy in predicting risk, as publicly available history-based risk algorithms such as the Gail model or the Tyrer Cuzick. When analyzing our miRNA data using elastic net, our average accuracy was 55%.

Thus, the premise upon which the rest of the project was to be built – that we can accurately classify cases of early onset breast cancer from controls on the basis of their RNA gene expression profile in the blood – did not hold up in our re-analysis. Thus, even though work had already begun on Specific Aim 3 – a validation cohort. We decided it best to discontinue this line of research (one year earlier than the end-date of the grant) in an effort to preserve the precious resources for a project with more promise.





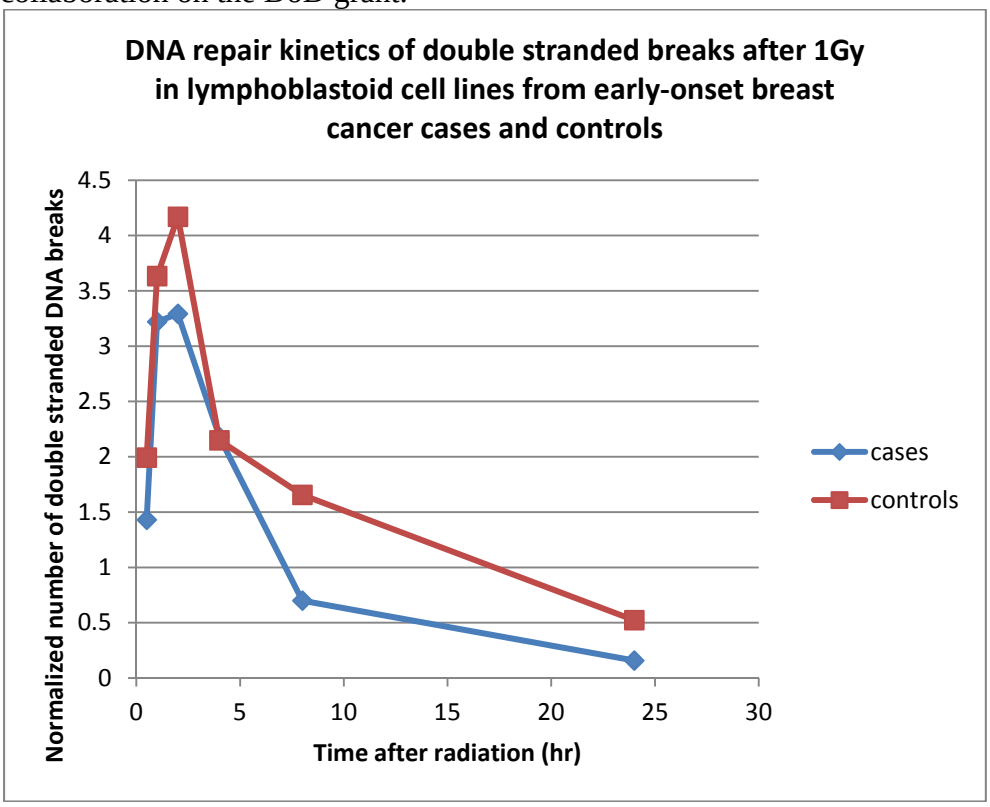
Specific Aim 2: To test whether a functional assay measuring DNA repair kinetics can accurately classify BRCA1/2 negative women with early onset breast cancer from age-matched asymptomatic women. (Months 7,8,9; 13-20)

Our initial aim was to compare the lymphoblastoid cell lines derived from the same cohort as in Aim 1, of 50 early-onset breast cancer cases and 51 controls, in their ability to repair DNA breaks using a unique assay developed by our collaborator, Dr. Sylvain Costes. We initiated a memorandum of understanding between UCSF and Lawrence National Berkeley labs (January 2014). We then provided cell lines in batches – equal numbers of cases and controls – and started growing them up. Unfortunately, we ran into difficulty on two fronts: 1. We discovered after submission of the grant, that in fact, we only had approximately half the number of lymphoblastoid lines than we believed were created initially. 2. Of these, only a fraction actually grew well in culture, so we are currently grossly underpowered.

We were able to get data on 6 cases and 5 controls, which are presented below.
The lymphoblastoid cell lines were subject to 1Gy of radiation exposure at timepoint zero, then the number of double-stranded DNA breaks was measured by the Costes Lab at 30 minutes, 1hr, 2hrs, 4hrs, 8hrs, and 24hrs in order to assess DNA repair kinetics (see figure below). We do not find any statistically significant differences between cases and controls at each timpoint (2 tailed t-tests), nor do we find any differences when comparing the delta between timepoint at maximal induction of DNA damage (1hr) and the 24hr timepoint (maximal repair), which would indicate degree of DNA repair.

In 2015 we again tried to re-grow some of these lymphoblastoid cell lines in culture in the lab of Dr. Sylvain Costes. Once again, the cells were quite sluggish in culture. It was decided to abandon this particular line of investigation. Despite the seeming failure of this particular

project/aim, our collaboration through the DoD grant led us to collaborating now on a clinical trials based project which is currently under review. This will be through the Sarah Cannon Research Network of which I am an affiliate, and Sylvain's Costes startup, Exogen Biotech. While outside the scope of the report of the DoD grant, the project's creation whose aim is to create a companion diagnostic based on DNA repair kinetics that can be used to better stratify women who are at high risk for breast cancer, and will rely on samples obtained in the clinical arena as part of routine high-risk screening workup, would not have been possible without our collaboration on the DoD grant.



Specific Aim 3: To validate the gene expression signature discovered in Specific Aims 1 and 2, in an independent prospectively collected cohort. Through collaboration with Drs. Michael Busch and Brian Custer at the Blood Systems Research Institute (BSRI), San Francisco, CA, we were able to access whole blood from 33 women with early onset breast cancer (onset before age 45) and 33 age-matched controls. This prospectively collected cohort consists of blood donated to blood banks ~15 years ago and subsequently linked to the California Cancer Registry. In this fashion, we have access to blood from women prior to the development of cancer. I have extracted total RNA on all of these samples (33 cases and 33 controls). In addition, 72 more cases of early onset breast cancer and 72 matched controls have been identified from the American Red Cross repository, with help from our collaborators at Blood Systems Research Institute. I have initiated the process of requesting access to these samples through NHLBI. However, once we realized that our initial analysis contained error, and that our re-analysis suggested that we could not accurately stratify early-onset cases from controls, we decided to halt the request of the additional samples from NHLBI which were to be used for validation. At

this time, it seems most prudent to discontinue further investigations related to this project's aims.

We set out to try to ascertain a gene-expression signature made up of mRNA and miRNA from a well annotated clinical cohort of women with early-onset breast cancer and age-matched asymptomatic controls. What we learned through the process was that we cannot come up with a gene signature or classifier that accurately stratifies these women into 'cases' and 'controls'. Thus far, commonly used history-based risk assessment algorithms such as the Gail model or the Tyrer Cuzick are just as effective. We still believe that DNA repair may be a powerful companion diagnostic to risk stratifying women deemed to be high risk by the aforementioned history-based risk algorithms and are the in process of designing and executing a clinical trial to test this hypothesis in the clinical arena. While outside the scope of this DoD postdoctoral Breast Cancer Research Fellowship award, it would not have been possible without this as a stepping stone. Additionally, during the time of the DoD award, I was able to contribute significantly (as second author) on a Nature Communications publication: Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25.

Opportunities for training and development: I applied for and was chosen to attend the Scientific Leadership and Management course held Fall 2014 at UCSF, modeled after that provided through HHMI.

How were the results disseminated to communities of interest: The results were presented locally within the UCSF community, as well as exernally to collaborators at the Blood Systems Research Institute and Illumina.

4. IMPACT: We set out to try to ascertain a gene-expression signature made up of mRNA and miRNA from a well annotated clinical cohort of women with early-onset breast cancer and age-matched asymptomatic controls. What we learned through the process was that we cannot come up with a gene signature or classifier that accurately stratifies these women into 'cases' and 'controls'. Thus far, commonly used history-based risk assessment algorithms such as the Gail model or the Tyrer Cuzick are just as effective. We still believe that DNA repair may be a powerful companion diagnostic to risk stratifying women deemed to be high risk by the aforementioned history-based risk algorithms and are the in process of designing and executing a clinical trial to test this hypothesis in the clinical arena. While outside the scope of this DoD postdoctoral Breast Cancer Research Fellowship award, it would not have been possible without this as a stepping stone. Additionally, during the time of the DoD award, I was able to contribute significantly (as second author) on a Nature Communications publication: Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25.

5. CHANGES/PROBLEMS: As outlined above, we had an error in our initial analysis which led us to believe we had a stronger classifier than we did in actuality. Once this was discovered, and after utilizing several alternate approaches to analysis of the same data, we realize that we are not able to build a reliable gene expression signature to accurately differentiate early-onset breast

cancer cases from controls. Hence, we feel it most prudent to discontinue further work on this particular project/grant at this time.

6. PRODUCTS: We had planned to centrally deposit the database of gene expression data once we published the results, to be accessible to all. However, as we have not and do not plan on publishing these "negative" results at this time, we have not deposited the data centrally.

I did have one second author publication, not directly related to this DoD grant, but while being supported by the DoD grant: "**Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25."** Nature Communications 2014 Oct 20; 5:5260. Fejerman L[1], **Ahmadiyeh N[2]**, Hu D[1], Huntsman S[1], Beckman KB[3], Caswell JL[1], Tsung K[2], John EM[4], Torres-Mejia G[5], Carvajal-Carmona L[6], Echeverry MM[7],Tuazon AM[7], Ramirez C[8]; COLUMBUS Consortium, Gignoux CR[9], Eng C[10], Gonzalez-Burchard E[10], Henderson B[11], Le Marchand L[12], Kooperberg C[13], Hou L[14], Agalliu I[15], Kraft P[16], Lindström S[16], Perez-Stable EJ[1], Haiman CA[11], Ziv E[1].

7. PARTICIPANTS and OTHER COLLABORATING ORGANIZATIONS:

Nasim Ahmadiyeh: PI (50% effort Year 1; 40% effort Year 2)

David Quigley: analyst (supported outside of the DoD grant)

Significant changes in active support of the PI or senior/key personnel: Nothing to Report

Partner Organizations:

Blood Systems Research Institute, San Francisco, CA – collaboration
Lawrence National Berkeley Laboratories, Berkeley, CA - collaboration

8. SPECIAL REPORTING REQUIREMENTS: none

9. APPENDICES: none